CS F425: Deep Learning

27 Siamese Network

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

April 06, 2023 ON-CAMPUS Campus @ BITS-Pilani [Jan-May 2023]

http://ktiwari.in/dl

Metric Learning

Learning in many applications depends upon **proximity measures**

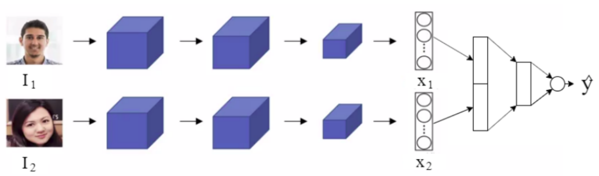
- Distance measure is used to find dissimilarity between pattern representations. More similar patterns should be near
- Metric "distance" measure $d(x, y) \geq 0$ has following properties
  1. *Positive reflexivity* $d(x, x) = 0$, therefore when $d(x, x) = 0 \rightarrow x = y$
  2. *Symmetry* $d(x, y) = d(y, x)$
  3. *Triangular Inequality* $d(x, y) \leq d(y, z) + d(z, y)$
- **Minkowski metric:**

$$d^m(X, Y) = \left(\sum_{k=1}^{d} |x_k - y_k|^m\right)^{\frac{1}{m}}$$

- Manhattan distance or L_1 uses $m = 1$, Euclidean or L_2 has $m = 2$
- L_0 is number of non-zero elements and L_∞ is maximum

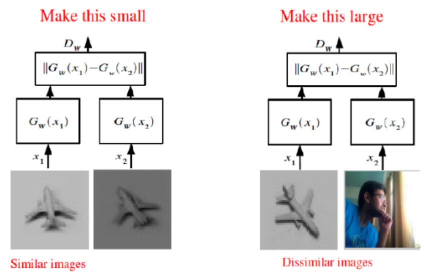
Siamese Network

Learn features using twin of an architecture.



- Weights are shared (same)
- Output could be the distance or classification
- Threshold could be used to determine if class is same or not?

Siamese Loss



- Contrastive loss

$$L = \sum \text{loss of positive pairs} + \sum \text{loss of negative pairs}$$

- Positive pairs $L(P, Q) = minimize||p - q||^2 \rightarrow EL$
- Negative pairs $L(P, Q) = maximize||p - q||^2$

Hinge Loss

- Maximization of $||p - q||^2$ is equavalant to minimization of

$$\infty - ||p - q||^2$$

- that is similar to minimization of $M - ||p - q||^2$
- Smaller value of M is used for multiclass setting

$$minimize(m - ||p - q||^2)$$

- However a **hinge** is better

$$L(P, Q) = max(0, m - ||p - q||^2) \rightarrow HL$$

Contrastive Loss that need to be minimized is

$$L = y \times EL + (1 - y) \times HL$$

Training: need of margin and the Triplet loss

Negative examples are much larger than positive ones.

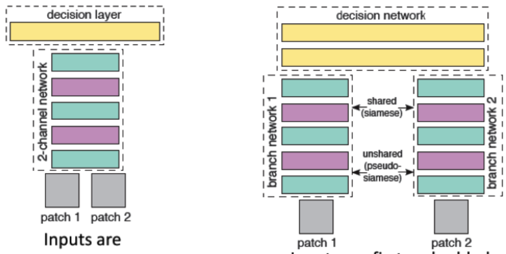
- Training would require a pair of images and 1/0
  With Anchore, Positive and Negative: (*A*,*P*,1), (*A*,*N*,0), (*P*,*N*,0), ...
- $D(A, P) = ||x_a - x_p||^2$ and $D(A, N) = ||x_a - x_n||^2$
  $D(A, P) \leq D(A, N)$ therefore we can have $D(A, P) + \alpha \leq D(A, N)$
- Or equivalently $\boxed{D(A, P) - D(A, N) + \alpha \leq 0}$
- Let

$$L(A, P, N) = max(D(A, P) - D(A, N) + \alpha, 0)$$

- Loss

$$J = \sum_{i=1}^{m} L(A^i, P^i, N^i)$$

Architecture Varieties



Inputs are merged right at the onset

Inputs are first embedded independently, then merged.

Design governs by what empirically performs well on the task at hand.

Online Semi Hard Negative Mining

- For a given network, we choose 1000 triplets randomly.
- Pass these 1000 triplets through the network and find D(A, P) and D(A, N) for each triplet.
- Select the triplets which satisfies the following conditions:
  - $D(A, N) - D(A, P) < \alpha$. These are the triplets which are hard and will produce a loss for the optimization of the network.
  - $D(A, N) > D(A, P)$. Discards the outliers.
- Out of the 1000 random triplets, we only train the network with those which satisfy above two conditions and then repeat this prosess.

Thank You!

Thank you very much for your attention!